

デジタルカメラワークを用いた自動映像生成

大西 正輝 泉 正夫 福永 邦雄
大阪府立大学 大学院工学研究科

Automatic Image Production Using Digital Camera Work

Masaki Onishi Masao Izumi Kunio Fukunaga
Graduate School of Engineering, Osaka Prefecture University

あらまし 近年、デジタル技術の発展はめざましく、情報通信分野の発展をもたらすものとして期待されている。中でも放送メディアのデジタル化は、多チャンネル・高画質映像などのサービスを提供できることから有望視されている。しかし、一方で番組製作側は数多くのコンテンツを用意する必要があることから、映像生成に十分な人手を費やすことが困難となり、映像生成の省力化が求められている。本稿では、カメラマンが撮影するときのカメラワークをコンピュータを用いた知的制御に基づいて実現することにより、自然な映像を自動的に生成する手法を提案する。まず、固定した一台のカメラを設置し、その映像から撮影要求の高い領域を切り出すことで構図とアングルを決定する。次にカメラを動かす理由を撮影の要求が変化する場合と、撮影対象として注目されるべき被写体が動く場合の2種類に分類してカメラワークを決定することで、カメラマンに近い撮影を行うことを可能にしている。また、体操の床演技を対象とした撮影規則を用いて実際に映像生成を行い、本手法の有効性を確認した。

キーワード 知的自動撮影, カメラワーク, 構図, アングル, カメラの動き, 体操の撮影

1 はじめに

デジタル多チャンネル時代を迎え、映像生成の省力化の要求が高まってきており、カメラマンが撮影するような自然な映像を自動で生成する研究が行われている。これらの技術は、スタジオ番組やスポーツ番組などの撮影だけではなく、災害地のような危険な場所やカメラマンが撮影できないような位置での映像生成も可能になるなど応用範囲は広い。

カメラマンはシーンを撮影する際にあらゆるシーン状況を理解し、経験に基づき最適なカメラポジションやカメラワークを決定しているものと考えられる。このようなカメラマンの柔軟な撮影知識を計算機に組み込み、自動的に最適な映像を撮ることを、ここでは知的自動撮影と呼ぶことにする。これまでも特定の撮影対象に対して知的自動撮影を行う手法が報告されており、講義を自動的に撮影する手法 [1, 2] や料理番組を自動的に撮影する手法 [3] などでは成果があげられている。しかし、これらの手法の多くはシーン中の状況理解を行い、その結果から撮影カメラの視線方向を決定する手法の提案であり、カメラの動きそのものについては述べられていない。このため、カメラの動きは目標とした視線方向に向かって一定の速度で動くか、あるいは瞬時に動くなどといった機械的な動きになってしまう可能性が高い。

プロのカメラマンによるカメラワークを分析する基礎研究として、台本であらかじめカメラワークが定まっているスタジオ番組や、スポーツ番組のように被写体の動きが予測できない状況下でのカメラワーク特性を調べる研究が行われており、その成果が報告されている [4, 5, 6]。本稿では、これらのカメラマンの撮影技術を模倣してリアルタイムで自然なカメラワークを実現する知的自動撮影手法を提案すると共に、体操の床演技を想定して作成した撮影規則を用いて自動撮影を行う。しかし、カメラマンがパンニングを行う際に生じるカメラの最大速度は 60 [deg/sec] 、最大加速度は $200 \text{ [deg/sec}^2\text{]}$ にも達するため [4]、市販の首振りカメラを制御するのは困難である。そこで、本稿では固定したカメラから撮影した映像中の一部分を切り取って、拡大呈示することで疑似的に映像を生成しており、これらの処理によって作り出すカメラワークをデジタルカメラワークと呼んでいる。デジタルカメラワークでは、切り取る領域を移動させることでパンニングやチルティング、切り取る領域の大きさを変えることでズーミングに相当する効果を生み出すことができる。

以下 2. で自然なカメラワークを作り出す手法について明らかにし、3. では体操の床演技を想定して撮影規則を作る方法を明らかにする。また、4. で実験・考察を行い、5. でまとめとする。

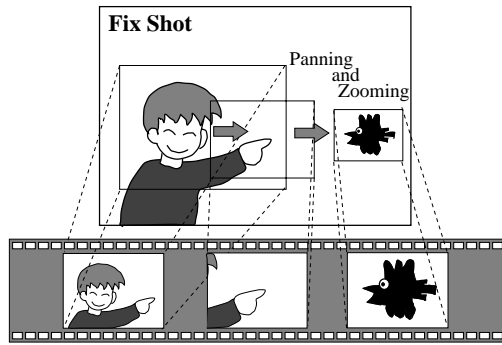


図 1: デジタルカメラワーク

2 カメラワーク

本研究では、固定カメラによって撮影した映像の一部分を切り取って拡大呈示することで映像の生成を行い疑似的にカメラを動かして撮影したかのように見せかけており、これらの処理をデジタルカメラワークと呼んでいる。デジタルカメラワークでは、切りとる領域を移動させることでパニングやチルディング、拡大率を変化させることでズームングの効果を作り出すことができる(図1)。また、生成される映像は固定カメラから得られた映像における切り取り位置の中心座標 x, y と拡大率 r によって決定できる。ここでは、これらの3つのパラメータ $x = (x, y, r)^T$ を撮影パラメータと呼ぶ。

カメラワークには構図、アングル、カメラの動きの三つの要素がある[7]。本手法のようにデジタルカメラワークを用いて固定カメラから得られた映像の一部分を切り出す映像生成法では、構図とアングルを決めることは撮影パラメータ x を決めることに相当し、カメラを動かすことは x を変化させることに相当する。以下、カメラワークの決定法について説明する。

2.1 構図とアングル

カメラマンは、構図やアングルを決定する際に“人物の顔を大きく撮影したい”、“人物の全身を撮影したい”、“背景を撮影したい”などといった様々な要求を満たすように撮影しているものと考えられる。そこで、本研究では様々な要求によって決定される撮影パラメータ(以下、撮影要求パラメータと呼ぶ)を幾つか設定しておき、それらの加重平均を求めることによって最終的な撮影パラメータを決定する。つまり、パラメータの要求数を n 、それぞれの撮影要求パラメータを x_i ($i = 1, 2, \dots, n$) として、撮影パラメータ x を

$$x = \sum_{i=1}^n A_i x_i \quad (1)$$

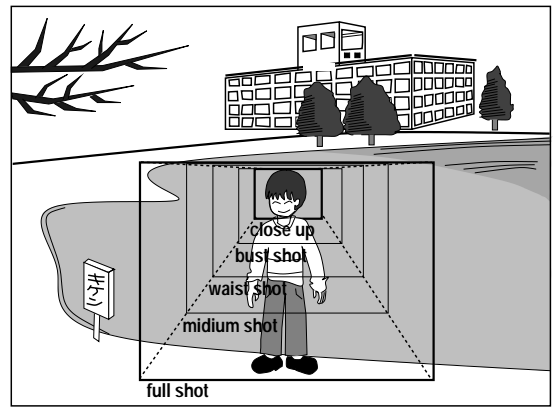


図 2: 人物の撮影

で決定する。行列 A_i は加重平均を取る時の重みを表し、対角成分以外は0、対角成分は0以上1以下の行列であり、

$$A_i = \begin{pmatrix} a_{11}^i & 0 & 0 \\ 0 & a_{22}^i & 0 \\ 0 & 0 & a_{33}^i \end{pmatrix} \quad (2)$$

とする。また、加重平均の重みであるため $e = (1, 1, 1)^T$ として、

$$\sum_{i=1}^n A_i e = e \quad (3)$$

を満たす。

本手法では撮影パラメータは幾つかの撮影要求パラメータの加重平均を求めることによって決定できると考えている。例えば、人物を撮影する際にクローズアップで顔を取りたいという要求とフルショットで全身を取りたいという要求があった場合に、これらの2つの要求に対する加重平均の重みを調整することでバストショット、ウェストショット、ミディアムショットなどを表現することが可能になる(図2)。また、これらの重みは、その時々によって決定されており、何かイベントが起こった場合に变化するものとする。

2.2 カメラの動き

デジタルカメラワークを用いて映像生成を行う場合、カメラを動かす作業は撮影パラメータ x を変化させることに相当する。式(1)では、カメラの動きを2種類に分類することができる。

一つは、移動しているものを撮影対象にしている場合で、例えば“歩いている人物を撮影したい”という要求がある場合などがこのケースに当たる。これは、式(1)において撮影要求パラメータ x_i が変化する場合に起こるカメラの動きに相当する。もう一つは、撮



図 3: 映像生成の例

影する対象を変える場合や対象が同一でも対象の見せ方を変える場合であり、撮影要求の重みが変化するために起こるカメラの動きである。例えば 2 人の人物 A, B がシーン中に存在しており、撮影要求が“人物 A を撮影したい”から“人物 B を撮影したい”に変化する場合には起こる。これは、式 (1) の重み A_i が変化することに相当する。前者のカメラの動きは、撮影要求パラメータの変化が引き起こすもので撮影要求パラメータの決定方法に依存しており、文献 [4] でも報告されている「サッカーの中継など被写体の動きが予測しにくい対象は、急激な加速や等速運動が見られ、規則性が見られない」という事実にも合致するカメラの動きであると考えられる。一方、後者のカメラの動きは撮影要求の重みの変化が引き起こすものであり、スタジオ番組のようにカメラワークがあらかじめ台本によって定まっている場合によく見られるカメラの動き [5, 6] であると考えられる。ここでは、後者のカメラの動きを実現する手法を説明する。

フレーム f_1 での重み $A_i^{f_1}$ をフレーム f_2 で重み $A_i^{f_2}$ に変化させる方法を考える。カメラを連続的に動かしているように見せるためには、重みを連続的に変化させる必要がある。そこで、フレーム f ($f_1 \leq f \leq f_2$) における重み A_i^f は、単位行列を E として、

$$A_i^f = \{E - B(\frac{f - f_1}{f_2 - f_1})\}A_i^{f_1} + B(\frac{f - f_1}{f_2 - f_1})A_i^{f_2} \quad (4)$$

で表す。ただし、

$$B(x) = \begin{pmatrix} b_{11}(x) & 0 & 0 \\ 0 & b_{22}(x) & 0 \\ 0 & 0 & b_{33}(x) \end{pmatrix} \quad (5)$$

とし、 $b(x)$ は単調増加する連続関数で、 $b(0) = 0$ 、 $b(1) = 1$ を満たす。 $f_2 - f_1$ が大きい程ゆっくりカメラを動かしているような効果を得ることができ、 $f_2 - f_1$ が小さい程カメラをはやく動かしているような効果が得られる。また、 $f_2 = f_1 + 1$ の時には瞬時に撮影パラメータが変わることになり、スイッチングの効果を得られる。

例えば $b(x) = x$ を用いた場合には、カメラが一定の速度で動いていることになり、機械的な動きに感じられるため視聴者は違和感を感じる。カメラワークに関する研究では一流カメラマンのパンニングやチルティング、ズームング時の速度曲線は非対称の三角波に近い形状であるとされている [5, 6]。そこで、関数 $b(x)$ は三角関数を用いた、

$$b(x) = \frac{1 - \cos(x\pi)}{2} \quad (6)$$

などを利用することで、実際にカメラマンがカメラを動かしているようなカメラワークが実現できると考えられる。

2.3 映像生成の実験

本手法を用いて実際の映像から映像生成を行った。固定カメラから得られる入力画像の解像度は 320×240 画素であり、デジタルカメラワークを用いて生成する画像は 160×120 画素とした。 $r = 2$ までは、解像度を落さずにズームすることができるが、 r が 2 以上の時にはデジタルズームを行った。また、シーン毎に重み A_i をあらかじめ決定しておき、イベントが起こった時に A_i を変化させる。ここでの実験は、本手法で生成したカメラワークに違和感がないかを調べることを目的としており、イベントが起こる時刻はあらかじめ撮影した映像を見て調べ、 A_i を変化させるタイミングや撮影要求パラメータの決定は手動で行った。

図 3 のシーン (a) は、映像中の人物が「黒板に文字を書き終える」と、黒板に書かれた文字にズームインするというシナリオで「黒板に文字を書き終える」というイベントが発生した時に板書文字をアップで映すという要求が大きくなるように撮影規則を作った。また、シーン (b) は、大学内の校舎全体の様子をパンで撮影していきパンの首振りが終わると校舎全体が映像中に収まるような映像に切り替えるというシナリオを設定した。シーン (a) は主にズームングの違和感を調べるために設定したシーンであり、シーン (b) は主にパンニング・スイッチングの違和感を調べるために設定したシーンである。図 3 の上段の映像が固定カメラから得られた映像であり、黒線の矩形が切り取り領域を示している。そして、これらの切り取った映像から下段のフィルム中の映像を作成した。本実験では、ほぼフレームレート (30 フレーム/秒) で処理を行っている。

下段の生成映像をビデオに撮影して、本研究の目的・内容を全く知らない被験者に呈示したところ、映像自体に違和感を感じたものはいなかったが、ズームイン時に解像度が低くなるという声があった。

3 撮影規則の作成

自動で映像生成を行うためには、撮影対象・目的に合わせた撮影規則を作る必要がある。本手法において、撮影規則を作ることは撮影要求パラメータの決定、イベントの決定、さらにはそのイベントが起こった時の重み A_i 、重みを変化させるための関数 $B(x)$ 、重みを変化させるのに必要なフレーム数 (f' とする) を決定することである。本節では、体操での床の演技を撮影対象に選び、撮影要求パラメータの決定法とイベント決定法を明らかにし、撮影規則を作る。

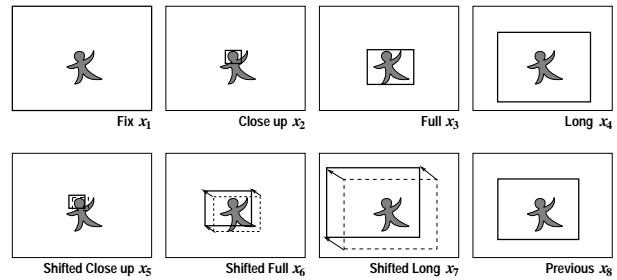


図 4: 構図とアングル

3.1 撮影要求パラメータの決定法

撮影対象は体操の床演技を想定しており、図 4 に示す 8 つの撮影要求パラメータを設定した。それぞれの決定方法について説明する。

- 固定映像全体を撮影したい (x_1) 固定カメラから取った映像全体を撮影したいという要求であり、 x_1 の x, y は固定カメラから得られる映像の中心、 r は 1 とする。
- クローズアップで撮影したい (x_2) クローズアップとは、人物の首から上を撮影することを言う。あらかじめ撮影しておいた背景画像と入力画像から差分画像を作り、差分値の大きい画素の色分布を調べ、肌色に近い画素をすべて含むように x_2 を定める。
- フルで撮影したい (x_3) フルとは、人物の全身を撮影することを言う。あらかじめ撮影しておいた背景画像と入力画像の差分画像を作り、差分値の大きい画素に接するように x_3 を定める。
- ロングで撮影したい (x_4) ロングとは人物だけではなく、周囲も映るように撮影することを言う。中心 x, y は、フルと同じものとして拡大率 r はフルの $1/2$ 倍にしたものとする。
- クローズアップに前あきを作りたい (x_5) 前あきとは、撮影する際に進行方向にあけるスペースのことである。前あきは、被写体のサイズや速度に依存することが報告されている。速度は過去 5 フレームにおいて差分値の高い画素の重心の平均移動量を用いて計算する。そして、前あきを作るために速度から x_2 の 5 フレーム先を線形予測したものを x_5 とする。
- フルに前あきを作りたい (x_6) フルの前あきは、速度から x_3 の 10 フレーム先を線形予測したものを x_6 とする。
- ロングに前あきを作りたい (x_7) ロングの前あきは、速度から x_4 の 15 フレーム先を線形予測したものを x_7 とする。
- 前と似た映像にしたい (x_8) 急激なカメラワークの変化は、視点が急に变化するなどの理由から望まれない。そこで、前フレームと似た映像を作る撮影要求

パラメータを用意する． x_8 は一つ前のフレームで用いた撮影パラメータ x を用いる．

最終的に生成映像を決定する撮影パラメータ x は、これら 8 つの撮影要求パラメータの加重平均によって式 (1) を用いて求める．

3.2 イベント 決定法

重み A_i は各イベントによって決定されており、イベントが検出された時点でそのイベントに合わせて重みが増える．多くのイベントを用意することで、多彩なカメラワークを実現することが可能になり、視聴者に興味を湧かせる映像になることが予想される．本稿では以下に示す 5 つのイベントを用意した．

- 選手が存在しない (Ev.1) 差分値の大きい画素が映像中にない場合を指す．シーン全体 (会場全体) を映す要求が高くなる．
- 選手が存在している (Ev.2) 差分値の大きい画素が映像中に存在する場合を指す．選手をズームインで映す要求が高くなる．
- 選手が速く移動している (Ev.3) Ev.2 を満たしており、差分値の大きい画素の重心の移動速度が一定速度以上の場合を指す．選手を映す時に前あきを大きく作って撮影する要求が高くなる．
- 選手がゆっくり移動している (Ev.4) Ev.2 を満たしており、差分値の大きい画素の重心の移動速度が 0 より大きく一定速度以下の場合を指す．選手を映す時に前あきはあまり必要ない．
- 選手が止まっている (Ev.5) Ev.2 を満たしており、差分値の大きい画素の重心の移動速度が 0 になる場合を指す．これは選手が技を決めた場合などに起こり、選手の表情を映すためにクローズアップで撮影する要求が高くなる．

また、これらのイベントが検出された時の撮影要求パラメータの重み A_i 、重みを変化させるのに必要なフレーム数 f' を表 1 に示す．本手法では、重み A_i の要素は $a_{11}^i = a_{22}^i = a_{33}^i$ としており、表中では a^i で表記している．これらの値は経験によって決めた．

表 1: 各イベントに対する撮影要求パラメータの重み

Ev.	a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8	f'
1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.9	30
2	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.8	30
3	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.8	30
4	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.8	30
5	0.0	0.2	0.1	0.0	0.1	0.0	0.0	0.6	1

4 実験・考察

4.1 実験

3. で説明した撮影要求パラメータとイベントを用いて、実際に自動で体操の撮影を行った．各イベントに対する重みは表 1 の値を用いる．また、 $b(x)$ は式 (6) を用いている．前述の実験と同様に固定カメラから得られる画像の解像度は 320×240 画素、生成映像の解像度は 160×120 画素とした．リアルタイムで撮影要求パラメータの算出やイベント抽出処理を行っているため、処理速度は 1 秒あたり 10 フレーム程度である．撮影映像の例を図 5 に示す¹．映像中に選手が登場するとズームインを行いロングショット程度で前あきを作りながら選手を追跡撮影して、技が決まった最後にはズームインして選手の表情を捉えていることが分かる．

4.2 考察

デジタルカメラワークを用いて自然な映像を自動で撮影する本手法について考察する．

- 自然なカメラワークを実現する手法の提案
スタジオ番組を撮影する場合のようにあらかじめ決まったカメラワークや、動きの予測できないものをより自然に撮影するためのカメラワークをコンピュータ制御によって実現するための手法を提案した．また、実際に撮影規則を作り本手法を用いて自動撮影を行い、本手法の有効性を確認した．実験では、固定カメラから得られた映像の一部分を切りとることで疑似的に映像生成を行っているが、カメラを実際に動かす場合にも本手法を適用することができると思われる．
- 撮影規則の作成
今回は撮影対象として体操の床演技を想定して、撮影要求パラメータを 8 つ、イベントを 5 つ作った．これらのパラメータやイベント数を多く設定することで多彩なカメラワークを実現することができ、視聴者に飽きさせない映像生成が可能になると思われる．
- 生成映像の評価
今回カメラを動かすために用いた重み A_i を変化させる関数は式 (6) を用いた．しかし実際のカメラマンのカメラワークは、このように数式で表現できるようなものではない．これらを数式で近似することによって視聴者にどのような感じを与えるかを評価していく必要がある．また、本手法に

¹<http://www.com.cs.osakafu-u.ac.jp/~onishi/research.html> を参照



図 5: 知的自動撮影による体操映像の生成

よって生成した映像をプロのカメラマンが撮影した映像と比較し、視聴者がどのような違いを感じるかを評価することが必要となるが今後の課題とする。

- 複数カメラによるスイッチングの問題

今回実験に用いた撮影カメラは一台であり、異なる視点からの映像は生成できない。撮影カメラを増やしてスイッチングを行うことで異なる視点からの映像を生成することや、瞬間的に映像の流れを作り出す効果を生むことが可能となる。しかし、スイッチングを行うためには複数のカメラから、より撮影映像として適した映像を選択する必要がある。さらに興味深い映像を生成するためには、効果的なスイッチング規則を反映できるように本手法を拡張する必要があるが、今後の課題とする。

- 解像度の問題

実験に用いた固定カメラから得られる映像の解像度は 320×240 画素であり、映像生成を行う際に十分な解像度で拡大呈示することができていない。近年 CCD カメラの研究は飛躍的に進歩しており、ハイビジョンカメラなどを用いて撮影した高解像度の映像を用いることで、鮮明な映像を得ることができると考えられる。

5 まとめ

本稿では、カメラマンが撮影するようなカメラワークを用いて、自動で映像を生成する手法を提案した。固定カメラから得られた画像から撮影要求の高い領域

を切り出し拡大呈示することで映像を生成する。また、体操の床演技を想定した撮影規則を作り、本手法を用いて人手が撮影したような映像を得ることができた。今後の課題として、本手法で生成した映像を評価することや、複数カメラでのスイッチング規則を考慮できるように手法を拡張することなどがあげられる。

参考文献

- [1] 大西正輝, 泉 正夫, 福永邦雄, “情報発生量の分布に基づく遠隔講義撮影の自動化,” 信学論 (D-II), vol. J82-D-II, no.10, pp.1590-1597, Oct. 1999.
- [2] 亀田能成, 石塚健太郎, 美濃導彦, “状況理解に基づく遠隔講義のための実時間映像化手法,” 情処学研報, 2000-CVIM-121-11, pp.81-88, Mar. 2000.
- [3] A.F.Bobick and C.S.Pinhanez, “Controlling View-Based Algorithms Using Approximate World Models and Action Information,” *Proc. Computer Vision and Pattern Recognition '97*, pp.955-961, June 1997.
- [4] 加藤大一郎, 石川秋男, 津田貴生, 福島 宏, 山田光穂, 阿部一雄, “動いている被写体を撮影時の放送カメラマンの撮影手法分析,” NHK 技研 R&D, no.48, pp.34-46, Feb. 1998.
- [5] 加藤大一郎, 山田光穂, 阿部一雄, “一流カメラマンのカメラワークと視線の動きの分析,” テレビ学技報, vol.19, no.6, pp.19-24, Feb. 1995.
- [6] 石川秋男, 加藤大一郎, 津田貴生, 福島 宏, 下田 茂, 山田光穂, 阿部一雄, “放送カメラマンのズーム計測法の検討と静止している被写体を撮影するときのズーム特性分析,” 映像メディア学会誌, vol.53, no.5, pp.749-757, May 1999.
- [7] 日本放送協会編, “NHK 撮影読本,” 日本放送出版協会, pp.48-56, 1980.